Available online at www.sciencedirect.com
SciVerse ScienceDirect
Procedia Engineering 15 (2011) 2328 – 2332
Procedia Engineering
www.elsevier.com/locate/procedia
Advanced in Control Engineeringand Information Science

A Special FCL Clustering and Its Application to Sparse Blind Source Separation

Yu Tong, Yunjie Zhang a*

Department of Mathematics, Dalian Maritime University, Dalian 116026, China

Abstract

Fuzzy *c*-Lines (FCL) algorithm is a linear fuzzy clustering algorithm and is constructed to treat linear dates and capture linear substructures. This paper considers a special FCL clustering algorithm, then based on the special FCL clustering algorithm and generalized inverse of vectors, proposes a new two-step clustering algorithms in order to solve the underdetermined sparse blind signal separation. The proposed algorithm provides a new approach for mixing matrix estimation and source signals separation, and simulation results support the validity of the algorithm.

Selection and/or peer-review under responsibility of [CEIS 2011]
Keywords: FCL; blind source separation; underdetermined sparse signal

1. Introduction

Fuzzy *c*-means (FCM) [1] is one of the well-known fuzzy clustering algorithm and many FCM modifications have been proposed after FCM. A first group of modifications is based on shapes of clusters. Linear fuzzy clustering algorithms such as Fuzzy *c*-Lines (FCL) [2] and Fuzzy *c*-Varieties (FCV) [3] were constructed to treat linear dates and capture linear substructures replacing the prototypical Fuzzy *c*-Means. Recent years, many people began to use Fuzzy clustering to deal with Blind Source Separation (BSS) [4, 5]. The objective of the BSS is to design a separation way to recover the original sources using only the information contained in observed signals. The simplest BSS model is the linear instantaneous model [6].

* Corresponding author. Tel.: 13478971929.
E-mail address: tongyu24@126.com.

doi:10.1016/j.proeng.2011.08.436

Suppose that $s(t) = (s_1(t), \ldots, s_n(t))^T$ is a n-dimensional source signal vector consisting of n source signals which are statistically independent each other. The sources are instantaneously and linearly mixed by an unknown mixing matrix $A = (a_{ij})_{m\times n}$ and observed as $x(t) = As(t)$, $t = 1, 2, \ldots, T$ or

$$x(t) = a_1 s_1(t) + \ldots + a_1 s_n(t),\ t = 1, 2, \ldots, T \tag{1}$$

where $x(t) = (x_1(t), \ldots, x_m(t))^T$ is an m-dimensional observed signal vector, and $a_j = (a_{1j}, \ldots, a_{mj})^T$ is A's jth column vector.

Most traditional BSS methods are independent component analysis (ICA) [7] and can't deal with underdetermined cases in which the number of observed signals is less than the number of source signals [8]. For the underdetermined BSS problem, researchers commonly restore all source signals by Sparse Component Analysis. Recently, two-stage approach is widely accepted to solve underdetermined sparse source separation [9-12]. Using two-stage approach, mixing matrix A is estimated by hyper-line clustering first, then source signals are recovered by the optimization strategic model. However, the clustering algorithms mentioned above are design for ellipsoid data, the method or process of solving the optimization model is very complex. Thus, it can hardly estimate mixture matrix accurately, and is rather computation intensive.

For this reason, this paper presents a new two-step clustering algorithm for underdetermined sparse blind signal separation. The proposed algorithm estimates mixing matrix using clustering algorithm via FCL Clustering, and reconstructs source signals via generalized inverse and inner product of vectors.

The remainder of this paper is organized as follows. In section 2, we give a special FCL Clustering algorithm. In section 3, we present a sparse source separation method using the special FCL Clustering given in 2. To illustrate the efficiency of algorithm, in the section 4, simulation results are given for blind separation of six sound signals. Conclusions are drawn in Section 5.

2. A Special FCL Clustering

In FCL clustering algorithm [2], the lines in R^S through a point v with unit direction vector d are denoted as $L(v; d) = \{y \in R^S \mid y = v + td;\ t \in R\}$. For sparse source separation, because the sensors signals will approximately form some lines through origin in m-dimensional geometric space, we can only consider a special FCL clustering algorithm: the lines in R^S are all through origin ($v = 0$), and they can be denoted as $L(d) = \{y \in R^S \mid y = td;\ t \in R\}$.

Supports data pairs in $S = \{(x_{1h}, \ldots, x_{mh})^T \mid h = 1, \ldots, N\}$ are drawn from c different hyper-lines: $L_i = \{x_h \in R^m \mid x_h = tb_i;\ t \in R\}$, $i = 1, \ldots, c$ or

$$\frac{x_{1h}}{b_{1i}} = \cdots = \frac{x_{mh}}{b_{mi}},\quad i = 1,\ \cdots,\ c \tag{2}$$

where $x_h = (x_{1h}, \ldots, x_{mh})^T \in R^m$ is variable and $b_i = (b_{1i}, \ldots, b_{mi})^T$ is the unit direction vector of L_i.

We can calculate the distance from $x_h = (x_{1h}, \ldots, x_{mh})^T$ to L_i as

$$d(x_h, L_i) = \| x_h - z^* \|^2 = \| ox_h \|^2 - \| oz^* \|^2 = \langle x_h, x_h \rangle - (\langle x_h, b_i \rangle)^2 \tag{3}$$

where o is origin, $z^*$is on the line L_i and $x_h z^* \perp L_i$.

Similar to the FCM, the objective function of the special FCL clustering algorithm is

$$J_m(U, b) = \sum_{h=1}^{N} \sum_{i=1}^{c} (u_{ih})^m (< x_h, x_h > -(< x_h, b_i >)^2) \tag{4}$$

where $m \in [1, +\infty)$ is weighting exponent.

Minimization of such an objective function in (4) yields simultaneous estimates for the directions of the c hyper-lines, together with a fuzzy c-partition of the data.

First, the same as FCM, we get fuzzy c-partition matrix U as

$$u_{ih}^{(r+1)} = \begin{cases} \left[\sum_{j=1}^{c} \left(d(x_h, b_i) / d(x_h, b_j) \right)^{2/(m-1)} \right]^{-1}, & I_h = \varnothing \\ 1/n_h, & I_h \neq \varnothing, i \in I_h \\ 0, & I_h \neq \varnothing, i \notin I_h \end{cases} \tag{5}$$

where $I_h = \{i \mid 1 \leq i \leq c, d(x_h, b_i) = 0\}$ and n_h is the number of elements in I_h. Second, for fixed U and x, minimizing the (4) is equivalent to maximize

$$\xi_i(b_i) = \sum_{h=1}^{N} (\hat{u}_{ih})^m (< b_i, x_h >< x_h, b_i >) = b_i^{\mathrm{T}} \left(\sum_{h=1}^{N} (\hat{u}_{ih})^m x_h x_h^{\mathrm{T}} \right) b_i, \quad i = 1, \cdots, c \tag{6}$$

That is that b_i (i = 1,…, c) is an eigenvector of

$$S_i = \sum_{h=1}^{N} (\hat{u}_{ih})^m x_h x_h^{\mathrm{T}}, \quad i = 1, \cdots, c \tag{7}$$

corresponding to its largest eigenvalue.

The FCL Clustering Algorithm is executed in the following steps:

Step1. Assign the number of the clusters c ($1 \leq c \leq n$), set the weighting exponent $m > 1$, specify the form of cluster representatives as (4), define $d(x_h, b_i)^2$ as (3), pick a termination threshold $\varepsilon > 0$ and an initial partition $U^{(0)}$. Set iteration index $r = 0$.

Step2. Calculate the c model parameters b_i: find the eigenvalue of (7) and the eigenvector b corresponding to the largest eigenvalue, set $b_i = b$.

Step3. Update $U^r \rightarrow U^{(r+1)}$ with $d(x_h, b_i)^2$ as (3).

Step4. Check for termination in some convenient induced matrix norms: If $\| U^r - U^{(r+1)} \| \leq \varepsilon$, stop; Otherwise, set $r = r + 1$ and return to Step 2.

3. Sparse Blind Source Separation Based on FCL Clustering

Generally, sparse signal is that whose most sample points are zero or are near to zero, and only small number points are far from zero [9]. So it is less possible for two source signals have large numbers in the same time, but only one sample point has a large number in almost all time.

Firstly, let $S = \{(x_1(t), \ldots, x_m(t)) \mid t = 1, \ldots, T\}$ is collection of observed signals and they are mixed by n sparse source signals. In the time of t, we support the source signal $s_j(t)$ is nonzero and the other source signals are zero or are near to zero, so the equation (1) can be written as:

$$x(t) = a_j s_j(t) \tag{8}$$

or

$$\frac{x_1(t)}{a_{1j}} = \cdots = \frac{x_m(t)}{a_{mj}} = s_j(t) \tag{9}$$

where $a_j = (a_{1j}, \ldots, a_{mj})^{\mathrm{T}}$, $x_1(t) = (x_1(t), \ldots, x_m(t))^{\mathrm{T}}$. All the sensors signals that the source signal $s_j(t)$ is nonzero will make a line that through origin and every $x(t)$ in the line is collinear with A's jth column vector a_j. Then all the data point in S approximately form n lines crossed through origin in m-dimensional geometric space. If we regard those $\{(x_1(t), \ldots, x_m(t))\}$ in the same line as one cluster, they come from the same linear equation as (9).

Secondly, jth (j = 1, …, n) cluster found by FCL clustering correspond with unique source signal $s_j(t)$, so from (8) we could get $s_j(t) = a_j^+ x(t)$ as jth source signal at t, where a_j^+ is generalized inverse of a_j. So, the sparse blind source separation algorithm based on FCL clustering is executed in follows:

Step1. Mark observable signals date set as date matrix $X = (x_{it})_{m\times T}$, where $x_{it} = x_i(t)$, m is the number of sensors signals, T is the samples number we take.

Step2. Given a number of cluster $c = n$ and construct n linear equations as (9).

Step3. Clustering for X based on FCL clustering, get linear equation parameter matrix $A = (a_{ij})_{m\times n}$ and fuzzy partition matrix $U = (u_{jt})_{n\times T}$ of S.

Step4. Let $A = (a_1, a_2, \ldots, a_n)$, and A is sought mixing matrix.

Step5. For $t = 1, \ldots, T$, find corresponding j such that $u_{jt} = \max\{u_{jt} \mid 1 \le j \le n\}$ and compute $s_j(t) = a_j^+ x(t)$.

Step6. For $t = 1, \ldots, T$, let $s(t) = (0, \ldots, 0, s_j(t), 0, \ldots, 0)^{\mathrm{T}}$, so $(s(1), s(2), \ldots, s(T))$ is source matrix (every row is one source signals in all the time).

4. Simulation Results

In order to illustrate the validity of new algorithm, we conducted six sound signals coming from www.ac.upe.es/homes/pau. The experimental analysis aims at comparing objectively the reconstruction index (Signal to Noise Rate) to measure performance of the proposed algorithm [12], and SNR is defined as:

$$S/N = -10\lg(\|\hat{s} - s\|^2 / \|s\|^2) \tag{10}$$

where $\hat{s}$ is the estimation of s. Both $\hat{s}$ and s are unitary signals, and have equal energy. The larger S/N is, the better the separation efficiency, and the validity of separation is good when $S/N \ge 25$.

Select 3, 4 and 5 respectively in 6 sound signals and full sparse, as the source signals, then randomly generated three mixing matrixes A_1, A_2, A_3 and got three group mixed signals.

Using the proposed sparse blind signal separation algorithm based on FCL clustering, we can get three estimated mixing matrixes W_1, W_2, W_3 and three group separated signals.

Here we only take the second group signals for example, four full sparse source signals are shown on Fig.1. (a), A_2 is the mixing matrixes, three sensors signals are shown on Fig.1. (b). Using the proposed sparse blind signal separation algorithm based on FCL clustering, we can get estimated mixing matrixes W_2 and four separated source signals which are shown on Fig.1.(c).

$$A_2 = \begin{bmatrix} 0.4155 & 0.7782 & 0.3818 & 0.2042 \\ 0.5910 & 0.5111 & 0.7122 & 0.4608 \\ 0.6915 & 0.3649 & 0.5890 & 0.8637 \end{bmatrix}, \quad W_2 = \begin{bmatrix} -0.4155 & -0.3818 & 0.2042 & -0.7782 \\ -0.5910 & -0.7122 & 0.4608 & -0.5111 \\ -0.6915 & -0.5890 & 0.8637 & -0.3649 \end{bmatrix}$$

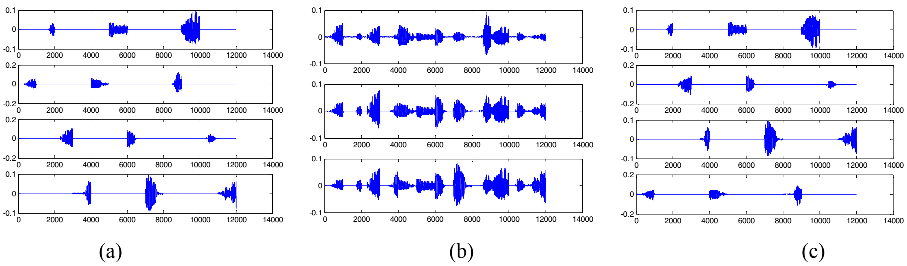

Fig. 1. The second group signals: (a) Four full sparse source signals; (b) Sensors signals mixed by four source signals; (c) Four separated source signals.

To analyze performance of algorithm proposed, we calculate reconstruction index (Signal to Noise Rate) for each separated signal using as (10) respectively. The evaluations are recorded in Table 1. From the results in Table 1, we can see that separation performance of algorithm is satisfying.

Table 1. Validity evaluation of three groups separation signals

	S/N of signal 1	*S/N* of signal 2	*S/N* of signal 3	*S/N* of signal 4	*S/N* of signal 5
Three source signals	707.7697	731.1494	722.6469		
Four source signals	726.1007	739.2460	730.8249	736.6981	
Five source signals	735.7387	728.9597	737.9557	728.3486	731.0818

5. Conclusion

Based on the special FCL Clustering algorithm and generalized inverse of vectors, this paper proposes a new two-step clustering algorithms in order to solve the underdetermined sparse blind signal separation. The proposed algorithm can effectively improve the mixing matrix estimation, and significantly reduce the costs of source signals recovery calculation. The simulation experiments illustrate the validity and some advantages of the proposed algorithm.

References

[1] F. Hoppner, F. Klawonn, R. Kruse and T. Runkler. *Fuzzy Cluster Analysis*. Jhon Wiley and Sons, 1999.

[2] J. C. Bezdek, C. Coray, R. Gunderson and J. Watson. Detection and characterization of cluster substructure I. Linear structure fuzzy c-lines. *SIAM Jour. of Appl. Math.* 1981; 40(2):339–357.

[3] J. C. Bezdek, C. Coray, R. Gunderson, and J. Watson. Detection and characterization of cluster substructure II. Fuzzy c-Varieties and convex combinations thereof. *SIAM Jour. Appl. Math.* 1981; 40(2):358–372.

[4] Owczarek AJ, *A new method of fuzzy clustering and its application to ECG signal analysis*. PhD Thesis. Silesian Technical University, Gliwice, 2004.

[5] B. H. Tan, Z. Y Yang, G. X. Z, Y. J Zhang. Source number determination and Separate Algorithm in Underdetermined Blind Separation, *Science in China Series F-Information Science*, 2009, Vol. 39, No. 3, 349~359.

[6] Haykin S. *Unsupervised Adaptive Filtering: Blind Source Separation*. New York, USA: Wiley J & Sons, 2000.

[7] J. Herault, C. Jutten. Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic. *Signal Processing*, 1991, 24(l):l-10.

[8] J. L. Zhang, S. L. Xie, Z. S. He. Separability theory for blind signal separation. *Acta Automatica Sinica*, 2004,vol. 30(3), pp. 337-344.

[9] Z. S. He, S. L. Xie, Y. L. Fu. Sparse representation and blind source separation of ill-posed mixtures. *Science in China Series F-Information Science*, 2006, vol.49 No.5pp.639-652.

[10] Y. Q. Li, A. Ciehocki, S. Amari. Analysis of sparse representation and blind source separation. *Neural computation*, 2004,vol. 16, no. 6, pp. 1193-1234.

[11] Z. S. He, A. Ciehocki. K-EVD Clustering and its Application in Sparse Component Analysis. *6th International Conference on Independent Component Analysis and Blind Signal Separation*, Charleston SC, USA, 2006, March 5-8.

[12] P. Bofill, M. Zibulevsky. Underdetermined source separation using sparse representation. *Signal Processing*, 2001, pp. 2353-2362.